**Author’s response to reviews**

**Title:** Potential gains in health expectancy by improving lifestyle: an application for European regions.

**Authors:**

Koen Füssenich (koen.fussenich@rivm.nl)

Wilma Nusselder (w.nusselder@erasmusmc.nl)

Stefan Lhachimi (stefan.lhachimi@uni-bremen.de)

Hendriek Boshuizen (hendriek.boshuizen@rivm.nl)

Talitha Feenstra (t.l.feenstra@umcg.nl)

**Version:** 2  **Date:** 29 Oct 2018

**Author’s response to reviews:**

Dear editor, dear reviewers,

We would like to thanks the reviewers for their time and effort to read our manuscript, but especially for the valuable comments made. In response, we hereby submit a revised manuscript, taking the comments made by the reviewers into account as much as possible. We have formulated a response for each of the comments listed below. The revisions to the manuscript are marked in red. We hope that with these changes, your concerns have been addressed adequately.

Kind regards,

On behalf of the authors,

Koen Füssenich

‘Reviewer reports:

Reviewer #1: I am happy that the reviewers have addressed all my comments.’

This reviewer has no further comments.
‘Reviewer #2: Dear Editor,

Dear Authors,

Thank you very much for sending the revised manuscript to me. Most of my criticisms have been well addressed. Nevertheless, I still see problems:

- The main issue I still have is the classification of BMI. I disagree with the authors who insist that they want to keep three categories. It is probably correct that there are very few data points for underweight people. But almost any research article, which addressed the issue of a BMI below 18.5 showed that mortality increases tremendously compared to people with "normal weight". Even if the estimates for underweight people are not-significant or problematic, using an extra category would avoid the distorted effect for "normal weight". My expectation would be that the differences between the categories would even become bigger as the baseline mortality risk of people with normal weight should decline once underweight persons are excluded from this category.’

We do agree with the reviewer that ideally we would separate ‘normal weight’ and ‘underweight’ due to the large difference in health. However, as the reviewer also mentions, applying this is very difficult due to the low number of underweight persons. We checked and in the Eurobarometer data, it is about 3% over all countries on aggregate. It will be difficult to derive any meaningful differences between countries based on these small numbers. Including an underweight category in the whole analysis is therefore very difficult.

However, the reviewer may have a valid point that including the underweight persons in the normal weight group might affect the observed differences in mortality between the normal and overweight group. Based on their prevalence, the underweight group will be about 5%-6% of the combined underweight and normal weight group. Any difference in health between normal and underweight group would have to be quite large to observe a strongly distorted effect.

Excluding the underweight group from all the analyses altogether might be a solution, however, this changes the composition of the population, on which other data, such as the number of newborns or the total mortality, is based. Hence we conclude that while theoretically valid, the probable effect on our results is relatively small. We have added a line in the discussion stating that our results for overweight might be an small underestimation of true effects, due to the inclusion of underweight persons in the normal weight category.
Furthermore, I can see that reviewer #1 has also major concerns with regard to alcohol related mortality. I was very happy to read that they addressed the "problems with alcohol" in more depth. But I have still this uncomfortable feeling how we can be sure that we can trust the estimates of the other risk factors? As far as I can see the whole package/software was only employed by the same people (checking the references and the DYNAMO-HIA website). Maybe this is the fate of many clever projects but it does not boost the confidence for outsiders.’

First, we should say the ‘problems with alcohol’ are largely unrelated to the DYNAMO-HIA model. The relative risk to mortality were used from the DYNAMO-HIA model, but these can be considered relatively reliable, as they are based on multiple other studies. The odds ratio’s and prevalence estimates however were derived from new data, the Eurobarometer, SHARE and HSM. The prevalence estimates are all from Eurobarometer and self reported. We mention in the discussion that this could lead to misreporting of smoking and BMI as well, however, upon face-value, they seem much more plausible that the numbers found for alcohol. We believe the number of glasses of alcohol is likely more easily misreported than smoking or BMI.

The same holds for the odds ratio’s found in SHARE. The figures found for smoking and BMI are corresponding much closer to previous literature (except smoking for females above 65), than those found for alcohol. Hence we added an analysis based on HSM to our SHARE analyses for alcohol. Since this database allowed more precise inclusion of age as a covariate and covered a broader age range, this helped to improve our estimates and we consider the resulting odds ratios as the best possible ones given data limitations.

The use of DYNAMO-HIA by only a small number of researchers is a different issue. While several researchers have made use of the model, often one of the original developers has co-authored most publications. Probably this is explained from the wish of adding in depth expertise of the model to the research team. However, examples exist of non involved researchers using the DYNAMO-HIA model (1-3).

‘-And what does it mean to select the best in "...and selected the best available to estimate risks" (line 107 of manuscript)?’

The relative risks to mortality are taken from the DYNAMO-HIA project and the collection is described in depth in the data documentation. We have removed the sentence as ‘selecting the best’ might not be the most accurate description of the study selection process, and refer to the data documentation for more in depth descriptions.
A minor issue is rather technical (and maybe personal curiosity). In the reply to my first set of comments, the authors write that they used the VGAM package of R. I mainly use the function gam() from package 'mgcv', which is included in any standard distribution of R, whereas VGAM is an add-on package. If gam() from 'mgcv' is not used, people use choose gam() from package 'gam', written by Trevor Hastie, one of the "inventors" or generalized additive models. So I wondered why the vgam package is used since the standard gam() can also estimate multinomial models. As I said: This is nothing crucial but would be interesting to me.'

VGAM is a package largely written by Thomas Yee, who has worked on VGLM's and VGAM's for many years, and also co-authored an article on this topic with Trevor Hastie. In the package description they briefly compare the VGAM package to the mgcv package, which might satisfy the technical curiosity in more depth. We did however choose to use the VGAM package due to previous experience within the institute with using VGAM for multinomial problems, and we did not purposely discard the mgcv package in favor of the VGAM package for this specific problem.

'Reviewer #3: The revisions of the text and presentation of results further improved the paper, which is, according to me, almost ready for publication. I have only a few additional minor suggestions:

Abstract: I suggest making background and conclusion better fitting to each other. In the current form, the conclusion focuses exclusively on the lagging health expectancy of Eastern Europe, while the background addresses the whole European region.'

We have changed the conclusion, decreasing the focus on Eastern Europe.

'Discussion (1): the section gained considerably with the additional arguments. However, these should be structured in a better way. For instance, the new text in lines 257-261 (p. 14) includes the same content as the new text in lines 274-278 (p. 15).'

While there was a small repetition about the underreporting in Eurobarometer, the sections touch upon slightly different issues with the alcohol data. First, there is the underreporting of heavy drinking. This is a prevalence estimate issue in the Eurobarometer data. The second part discusses the relation between drinking and LEGPH and HLY. Does drinking more lead to worse
health outcomes? This is a largely different issue from the amount of heavy drinkers. We have taken out the repetition about the underreporting.

‘Discussion (2): in my first review, I suggested to mention that the smaller smoking effects among women might increase in the future because of the lagged - and in most countries still increasing - smoking prevalence compared to men. The authors rejected this suggestion because they "do not think that smoking in women will ever reach the levels of exposure experienced by men in the past". I also do not believe in identical peak values among women and men, and this was not the point of my argument. Unfortunately, it happens very often that readers interpret cross-sectional results in a longitudinal or forecasting manner. This is why authors should make the features of cross-sectional analyses as clear as possible. The authors of this paper extended the discussion by these aspects, what I highly appreciate. However, in one of the new text parts they explicitly describe that the decreasing smoking prevalence among men should lead to a lower potential effect of smoking reduction in the future (p. 15, lines 292-295). Thus, I urge the authors to reconsider their rejection of my suggestion to mention also the different situation for women to avoid a misunderstanding. The current text might be understood in the way that also among women a reduced potential effect of smoking reduction should be expected for the future, what is not in line with the past and actual trends in female smoking prevalence.’

We would like to thank the reviewer for the clarification of the comment. We have taken the suggestion into account and added a line separating the trend for males and females to hopefully prevent misunderstandings.

References

